Powell eBook Prediction Models

An analysis of customer data to identify key factors influencing 'Average Monthly Spending' and potential 'eBook subscribers' within a customer base.

Key Customer Characteristics

Family Structure & Spending

Primary influencers on eBook subscriptions.

Income & Family Size

Also play a significant role in subscription likelihood.

Marital Status, Occupation, Age

Moderate impact on eBook subscription decisions.

Understanding these drivers informs targeted marketing and engagement strategies for eBook subscribers.

Predictors of Monthly Spend

Children at Home

Strongest predictor of average monthly spend.

Annual Income

Significant correlation with spending habits.

Gender

Another key factor influencing monthly spend.

Total Children

Contributes to predicting spending behavior.

Occupation

Moderate influence on average monthly spend.

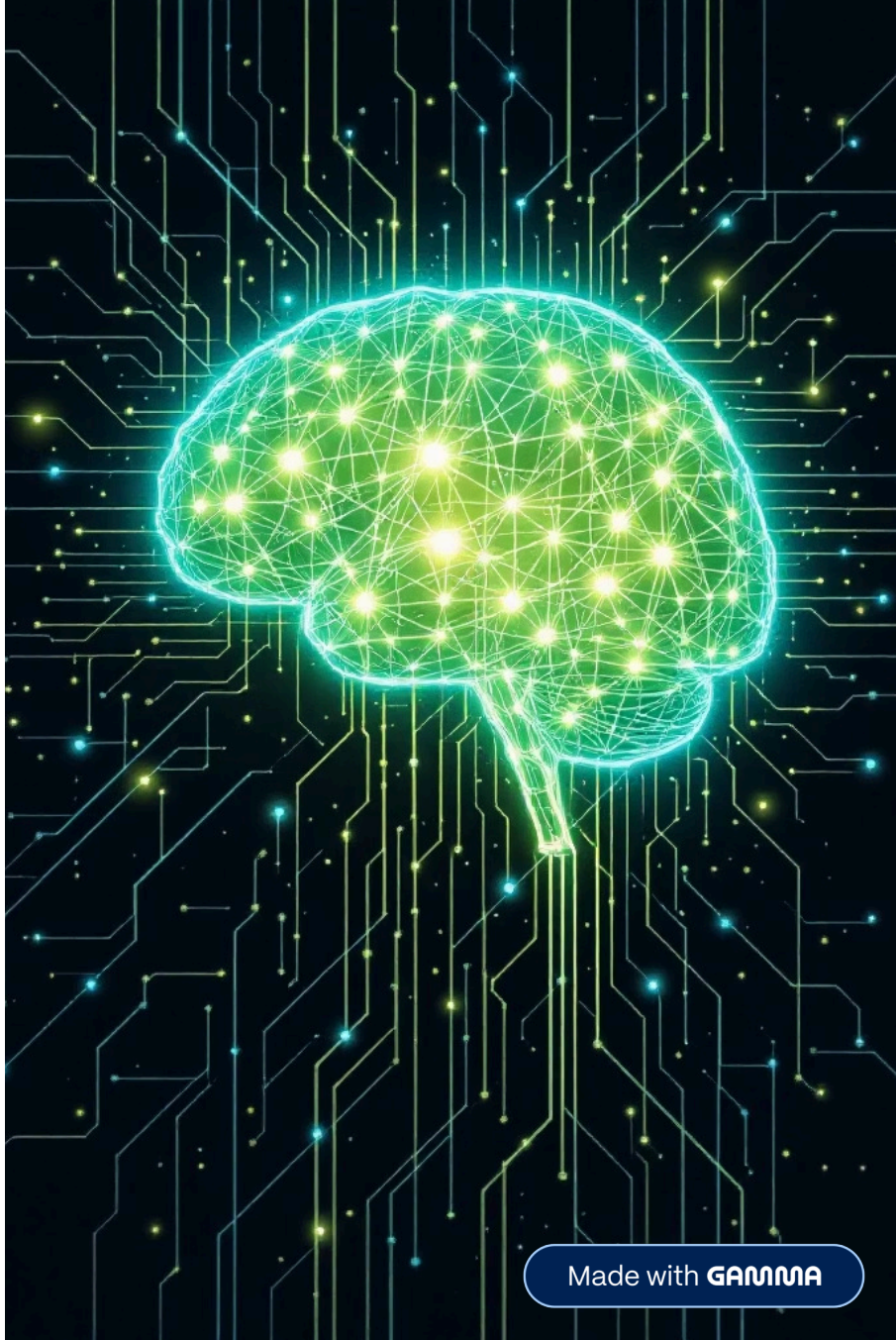
Cars Owned

Also a factor in predicting spending.

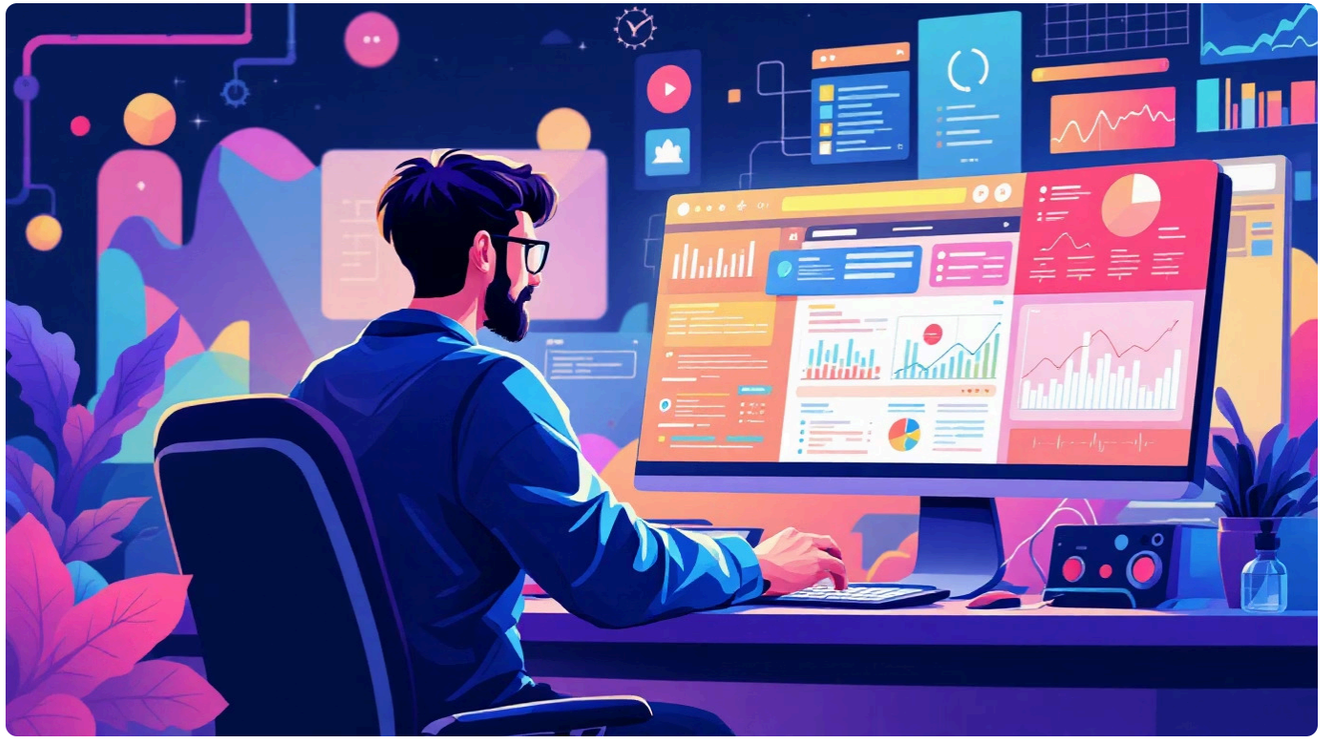
These factors, in descending order of influence, are crucial for tailoring product offerings.

Developing Prediction Models

PyCaret: A machine learning model was developed to accurately predict 'Average Monthly Spending' and potential 'eBook subscribers'. This provides insights for targeted marketing and customer engagement strategies.



Model Development: Data Preparation



01

Load Dataset

16,519 records across 25 variables were loaded.

02

Feature Engineering

'Age' was engineered from 'Birth Date'.

03

Parse Location Data

'City-ZipCode-State' parsed into distinct fields.

04

Exclude Irrelevant Columns

Identified and removed missing and redundant data.

Model Development: Setup & Training

01

Set Up Models

Two models (classification and regression) were set up using PyCaret.

02

Define Target Variables

'eBook Subscriber Flag' and 'Avg Monthly Spending' were specified.

03

Specify Features

Relevant categorical and numerical features were defined.

Model Selection & Optimization

Compare Algorithms

PyCaret's `compare_models()` evaluated various algorithms.

Subscriber Prediction

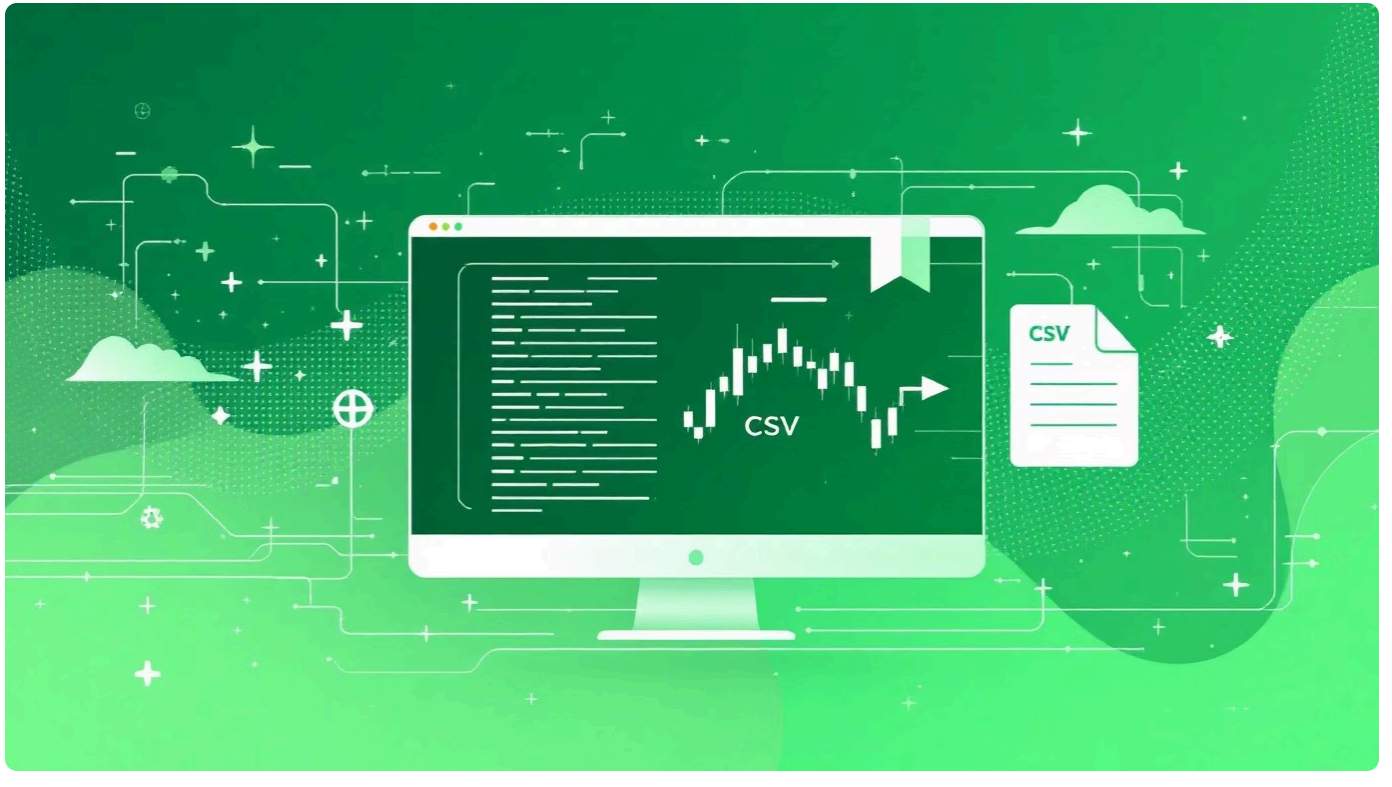
Gradient Boosting Classifier was selected for subscriber prediction.

Spend Prediction

Light Gradient Boosting Machine (LightGBM) was selected for monthly spend.

This rigorous selection process ensured optimal model performance for both predictions.

Prediction & Output



Apply Optimized Models

Models applied to the dataset for individual customer predictions.

Generate Predictions

Predictions for 'eBook Subscriber Flag' and 'Avg Monthly Spending' were generated.

Export to CSV

Predictions exported to CSV files for further use.

Key Findings Summary

Our analysis reveals a robust predictive model for customer spending behavior, providing actionable insights for strategic planning.

High Accuracy

98.63% R-squared indicates the model explains nearly all variance in monthly spending.

Spending Prediction

Accurately predicts average monthly spending at **$72.41** per customer.

Extensive Data

Analysis based on a comprehensive dataset of **16,519 customers**.

Business Ready

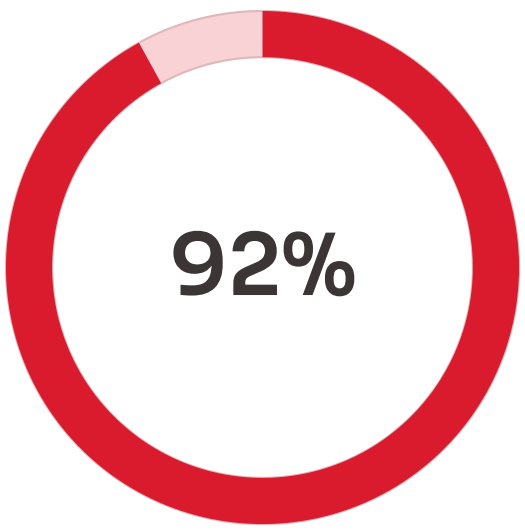
The model is ready for immediate application in targeted marketing and resource allocation.

View Full Analysis & Data

Key Findings Summary

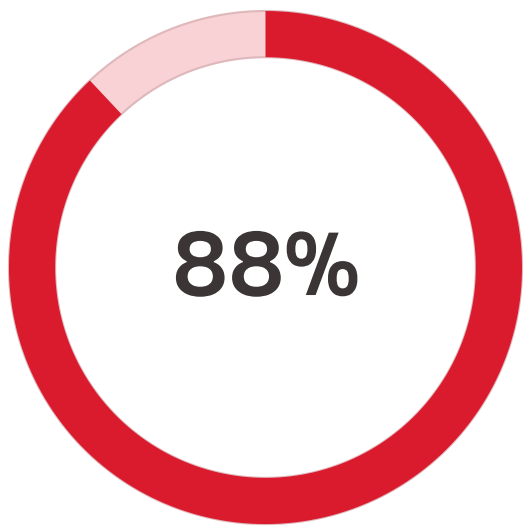
Our analysis successfully identified key patterns and predicted eBook subscription status with high confidence, providing actionable insights for strategic decision-making.

Model Performance & Customer Segmentation



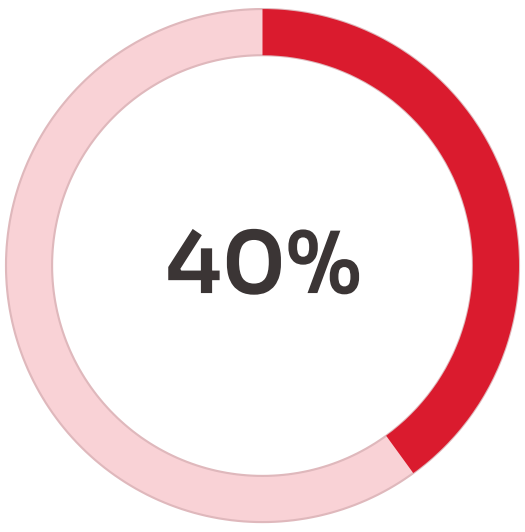
Prediction Accuracy

The Gradient Boosting Classifier achieved strong overall accuracy.



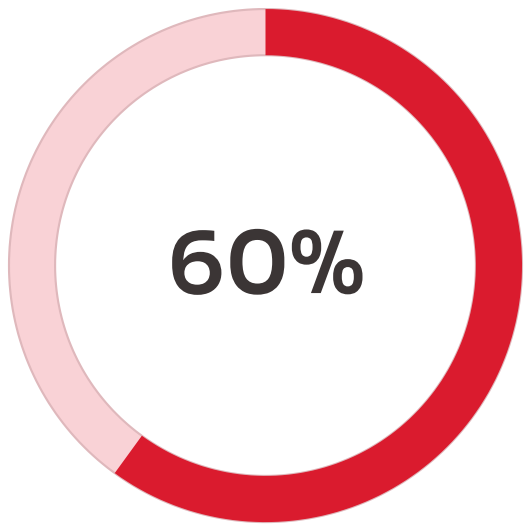
F1-Score

Robust performance across both subscriber and non-subscriber classes.



Predicted Subscribers

Identified proportion of the dataset likely to be eBook subscribers.



Predicted Non-Subscribers

Segment of customers less likely to engage with eBook subscriptions.

Top Factors Influencing Subscription Status

Annual Income

Higher income levels showed a positive correlation with subscription.

Age Group

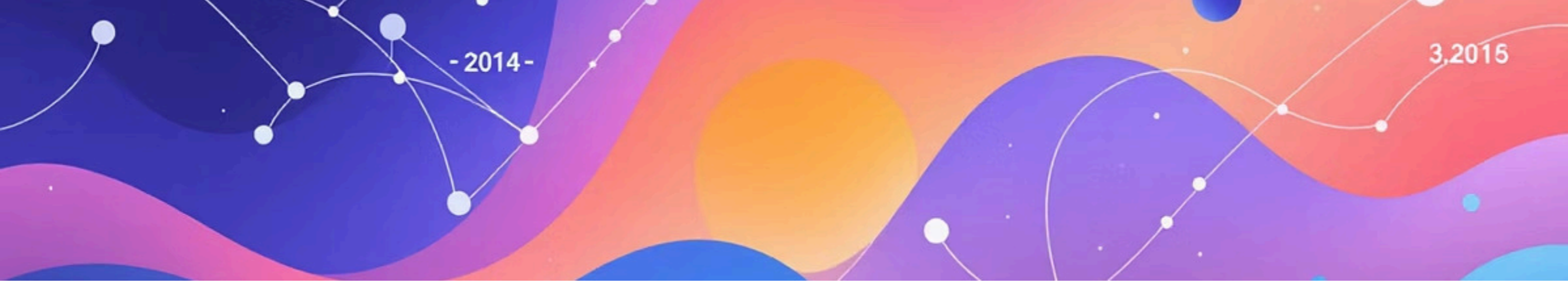
Customers in the 25-45 age bracket were more likely to subscribe.

Home Ownership

Homeowners exhibited a slightly higher propensity for subscriptions.

Occupation

Specific professional occupations were strong indicators of subscription interest.



Comprehensive Summary

This comprehensive analysis delves into two critical prediction models: one designed to identify the likelihood of existing customers subscribing to an eBook service, and another focused on forecasting the average monthly spending patterns of customers. Together, these models provide a dual perspective on customer behavior, allowing for more nuanced strategic planning.

Critical Finding: Variables Impact Both Predictions Differently

While factors like **family structure, income levels, age demographics,** and **occupational details** are key predictors for both eBook subscription likelihood and monthly spending, their specific influence and weighting vary significantly between the models. This suggests a nuanced relationship where the same variable can drive distinct behavioral outcomes.

Methodology

Our approach leverages advanced machine learning techniques, specifically utilizing the **PyCaret library** for streamlined model development and deployment. We analyzed a robust dataset comprising **16,519 customer records** enriched with **25 distinct variables**.

- **Dual-model strategy:** A Gradient Boosting Classifier was employed for predicting eBook subscription likelihood.
- **Dual-model strategy:** A LightGBM model was developed for forecasting average monthly spending.

Business Value

The insights from these prediction models hold substantial value for Powell, empowering data-driven decisions:

- Develop **highly targeted marketing strategies** by understanding specific customer segments.
- Segment customers more effectively for **personalized engagements**.
- Optimize **resource allocation** by focusing efforts on customers most likely to convert.
- Gain a proactive understanding of **future revenue potential**.